pKa measurements for the SAMPL6 prediction challenge for a set of kinase inhibitor-like fragments

Mehtap Işık[1,2], Dorothy Levorse[3], Ariën S. Rustenburg[1,4], Ikenna E. Ndukwe[5], Heather Wang[6], Xiao Wang[5], Mikhail Reibarkh[5], Gary E. Martin[5], David Mobley[7], Timothy Rhodes[3], John D. Chodera[1*]

[1]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States; [2]Tri-Institutional PhD Program in Chemical Biology, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065, United States; [3]Merck & Co., Inc., MRL, Pharmaceutical Sciences, 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States; [4]Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY 10065, United States; [5]Merck & Co., Inc., MRL, NMR Structure Elucidation, 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States; [6]Merck & Co., Inc., MRL, Process Research & Development, 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States; [7]Department of Pharmaceutical Sciences and Department of Chemistry, University of California, Irvine, Irvine, California 92697, United States

***For correspondence:**
john.chodera@choderalab.org (JDC)

Abstract Determining the protonation states populated by a small molecule in an environment of interest—such as solvent, a protein binding site, or a lipid bilayer—is a prerequisite for predicting its physicochemical and pharmaceutical properties, as well as interactions with biological macromolecules using computational models. Incorrectly modeling the dominant protonation state, shifts in dominant protonation state, or the population of significant mixtures of protonation states can lead to large modeling error that degrade the accuracy of physical modeling and hinder the ability to use physical modeling approaches for molecular design. For small molecules with a single titratable group, the acid dissociation constant (pKa) is the primary quantity needed to determine the populations of ionic states populated by a molecule in an aqueous solution at a given pH. As a part of SAMPL6 community challenge, we organized a blind pKa prediction component to assess the accuracy with which contemporary pKa prediction methods can predict this quantity, with the ultimate aim of assessing the expected impact on modeling errors this would induce. While a multitude of approaches to predicting pKas currently exist, predicting the pKas of drug-like molecules can be difficult due to challenging properties such as multiple titratable sites, heterocycles, and tautomerization. For this challenge, we focused on set of 24 small molecules selected to resemble selective kinase inhibitors—an important class of therapeutics replete with titratable moieties. Using a Sirius T3 instrument that performs automated acid-base titrations, we used UV-absorbance based pKa measurements to construct a high-quality experimental reference dataset of macroscopic pKas for the evaluation of computational pKa prediction methodologies that was utilized in the SAMPL6 pKa challenge. For several compounds in which the microscopic protonation states associated with macroscopic pKas were ambiguous, we performed follow-up NMR experiments to disambiguate the microstates involved in the transition. This dataset provides a useful standard benchmark dataset for the evaluation of pKa prediction methodologies on kinase inhibitor-like compounds.

Keywords
acid dissociation constants · spectrophotometric pKa measurement · blind prediction challenge · SAMPL · macroscopic pKa · microscopic pKa · macroscopic protonation state · microscopic protonation state

Abbreviations
SAMPL Statistical Assessment of the Modeling of Proteins and Ligands
pKa $-log_{10}$ acid dissociation equilibrium constant
psKa $-log_{10}$ apparent acid dissociation equilibrium constant in cosolvent
DMSO Dimethyl sulfoxide
ISA Ionic-strength adjusted
SEM Standard error of the mean
TFA Target factor analysis
LC-MS Liquid chromatography - mass spectrometry
NMR Nuclear magnetic resonance spectroscopy

Introduction

SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) is a recurring series of blind prediction challenges for the computational chemistry community [1, 2]. Through these challenges, SAMPL aims to evaluate and advance computational tools for rational drug design: By focusing the community on specific phenomena relevant to drug discovery poorly predicted by current models, isolating that phenomenon from other confounding factors in well-designed test systems, evaluating tools prospectively, enforcing data sharing to learn from failures, and releasing the resulting high-quality datasets into the community as benchmark sets, SAMPL has driven progress in a number of areas over seven previous rounds of challenge cycles.

JDC: Cite SAMPL overview papers.

As a stepping stone to enabling the accurate prediction of protein-ligand binding affinities, SAMPL has focused on evaluating how well physical and empirical modeling methodologies can predict various physicochemical properties relevant to binding and drug discovery, such as hydration free energies (which model aspects of desolvation in isolation), distribution coefficients (which models transfer from aqueous to nonpolar environments without conflating slow protein dynamics), and host-guest binding affinities (which models high-affinity association without the complication of slow protein dynamics). These physicochemical property prediction challenges—in addition to assessing predictive accuracy of quantities that are useful in various stages of drug discovery in their own right—have been helpful in pinpoint deficiencies in computational models that can lead to substantial errors in affinity predictions.

JDC: Cite some SAMPL papers highlighting lessons learned.

Neglect of protonation state effects can lead to large modeling errors

As part of the SAMPL5 challenge series, a new cyclohexane-water distribution constant (logD) prediction challenge was introduced, where participants predicted the transfer free energy of small drug-like molecules between an aqueous buffer phase at pH 7.4 and a nonaqueous cyclohexane phase [3, 4]. While octanol-water distribution coefficient measurements are more common, cyclohexane was selected for the simplicity of its liquid phase and relative dryness compared to wet octanaol phases. While the expectation was that this challenge would be relatively straightforward given the lack of complexity of cyclohexane phases, analysis of participant performance revealed that multiple factors contributed to significant prediction failures: poor conformational sampling of flexible solute molecules, misprediction of relevant protonation and tautomeric states (or failure to accommodate shifts in their populations), and force field inaccuracies resulting in bias towards the cyclohexane phase. While these findings justified the benefit of future iterations of blind distribution or partition coefficient challenges, the most surprising observation from this initial logD challenge was that participants almost uniformly neglected to accurately model protonation state effects, and that neglect of these effects led to surprisingly large errors in transfer free energies [3–5]. Careful quantum chemical assessments of the magnitude of these protonation state effects found that their

neglect could introduce errors up to 6–8 kcal/mol for some compounds [5]. To isolate this surprisingly large protonation state effects modeling error from difficulties related to lipophilicity (logP and logD) prediction methods, we decided to organize a set of staged physicochemical property challenges using a consistent set of molecules that resemble small molecule kinase inhibitors—an important drug class replete with titratable moieties. This series of challenges will first evaluate the ability of current-generation modeling tools to predict acid dissociation constants (pKa), followed by the ability to incorporate experimentally-provided pKas into prediction of distribution coefficients to ensure methodologies can correctly incorporate protonation state effects into their predictions, followed by a new blinded partition coefficient challenge where participants must predict pKa values on their own. At the conclusion of this series of challenges, we will have ensured that modern physical and empirical modeling methods have eliminated this large source of spurious errors from modeling both simple and complex phenomena. This article reports the first stage of this series of challenges: The selection of a small molecule set and collection of experimental pKa data for the SAMPL6 pKa challenge.

Conceptualization of a blind pKa challenge

This is the first time a blind pKa prediction challenge has been fielded as part of SAMPL. In this first iteration of the challenge, we aimed to assess the performance of current pKa prediction methods and isolate potential causes of inaccurate pKa estimates, with the aim of determining how pKa prediction inaccuracies might impact predicted affinities for drug-like molecules. For example, for both logD and binding affinity predictions, any error in predicting the free energy of accessing a minor protonation state in solution that becomes dominant in the complex will directly add to the error in the predicted transfer or binding free energy.

The prediction of pKas for drug-like molecules can be complicated by several effects: the presence of multiple (potentially coupled) titratable sites, the presence of heterocycles, tautomerization, the conformational flexibility of large molecules, and ability of intramolecular hydrogen bonds to form. We decided to focus on the chemical space of small molecule kinase inhibitors in the first iteration of pKa prediction challenge. A total of 24 small organic molecules (17 fragment-like and 7 drug-like) were selected for their similarity to known small molecule kinase inhibitors, while also considering properties predicted to affect the experimental tractability of pKa and logP measurements such as solubility and predicted pKas. Macroscopic pKa values were collected experimentally with UV-absorbance spectroscopy based pKa measurements using a Sirius T3 instrument, which automates the sample handling, titration, and spectroscopic measurement process to allow high-quality measurements on the order of tens of compounds. Experimental data was kept blinded for three months (25 Oct 2017 through 23 Jan 2018) to allow participants in the SAMPL6 pKa challenge to submit truly blinded computational predictions. Eleven research groups participated in this challenge, providing a total of 93 prediction submission sets that cover a large variety of contemporary pKa prediction methods.

Our selected experimental approach determines macroscopic pKa values

Whenever experimental pKa measurements are used for evaluating pKa predictions, it is important to differentiate between microscopic and macroscopic pKa values. In molecules containing multiple titratable moieties, the protonation state of one group can affect the proton dissociation propensity of another functional group. In such cases, the **microscopic pKa** refers to the pKa of deprotonation of a single titratable group while all the other titratable and tautomerizable functional groups of the same molecule are held fixed. Different protonation states and tautomer combinations constitute different microstates. The **macroscopic pKa** defines the acid dissociation constant related to the observable loss of a proton from a molecule regardless of which functional group the proton is dissociating from, so it doesn't necessarily convey structural information.

Whether a measured pKa is microscopic or macroscopic depends on the experimental method used (Figure 1). For a molecule with only one titratable proton, the microscopic pKa is necessarily equal to the macroscopic pKa. For a molecule with multiple titratable groups, however, throughout a titration from acidic to basic pH, the deprotonation of some functional groups can take place almost simultaneously.

For these multiprotic molecules, the experimentally-measured macroscopic pKa will include contributions from multiple microscopic pKas with similar values (i.e., acid dissociation of multiple microstates). Cysteine provides an example of this behavior with its two macroscopic pKas observable by spectrophotometric or potentiometric pKa measurement experiments [6]. While four microscopic pKas can be defined for cysteine, experimentally observed pKas can't be assigned to individual functional groups directly, and more advanced techniques capable of resolving individual protonation sites—such as NMR [7], Raman spectroscopy [8, 9], and analysis of pKas in molecular fragments or derivatives—are required to unambiguously assign protonation state changes. On the other hand, when there is a large difference between microscopic pKas in a multiprotic molecule, the proton dissociations won't overlap and macroscopic pKas observed by experiments can be assigned to individual titratable groups. The pKa values of glycine provide a good example of this scenario [6, 8]. We recommend a short review on the assignment of pKa values authored by Ivan G. Darvey. [6] for a good introduction to the concepts of macroscopic vs. microscopic pKas.

UV-absorbance spectroscopy (UV-metric titration) [12–14], potentiometry (pH-metric titration) [14, 15], capillary electrophoresis [16, 17], and NMR spectroscopy [7] are the most common methods for measuring small molecule pKas. Other, less popular pKa measurement techniques include conductometry, HPLC, solubility or partition based estimations, calorimetry, fluorometry, and polarimetry [18]. UV-metric and pH-metric methods (Figure 2) are limited to measuring aqueous pKa values between 2 and 12 due to limitations of the pH-electrode used in these measurements. The pH-metric method relies on determining the stochiometry of bound protons with respect to pH, calculated from volumetric titration with acid or base solutions. Accurate pH-metric measurements require high concentrations of analyte and analytically prepared acid/base stocks and analyte solutions. By contrast, UV-metric pKa measurements rely on the differences in UV-absorbance spectra of different protonation states, generally permitting lower concentrations of analyte to be used. pH and UV absorbance of analyte solution is monitored during titration.

Both UV-metric and pH-metric pKa determination methods measure macroscopic pKas for polyprotic molecules, which can not be easily assigned to individual titration sites and underlying microstate populations in the absence of other experimental evidence that provides structural information, such as NMR (Figure 1). Macroscopic populations observed in these two methods are composed of different combinations of microstates depending on the principles of measurement technique. In potentiometric titrations, microstates with same total charge will be observed as one macrostate, while in spectrophotometric titrations, protonation sites remote from chromophores might be spectroscopically invisible, and macrostates will be formed from collections of microstates that manifest similar UV-absorbance spectra.

Spectrophotometric pKa determination is more sensitive than potentiometric determination, requiring low analyte concentrations (50–100 µM)—especially advantageous for compounds with low solubilities—but is only applicable to titration sites near chromophores. For protonation state changes to affect UV absorbance, a useful rule of thumb is that the protonation site should be a maximum of four heavy atoms away from the chromophore, which might consist of conjugated double bonds, carbonyl groups, aromatic rings, etc. Although potentiometric measurements do not suffer from the same observability limitations, higher analyte concentrations (~5 mM) are necessary for the analyte to provide sufficiently large enough buffering capacity signal above water to produce an accurate measurement. The accuracy of pKas fit to potentiometric titrations can also be also sensitive to errors in the estimated concentration of the analyte in the sample solution, while UV-metric titrations are insensitive to concentration errors. We therefore decided to adopt spectrophotometric measurements for collecting the experimental pKa data for this challenge, and selected a compound set to ensure that all potential titration sites are in the vicinity of UV-chromophores.

Here, we report on the selection of SAMPL6 pKa challenge compounds, their macroscopic pKa values measured by UV-metric titrations using a Sirius T3, and NMR-based microstate characterization of two SAMPL6 compounds with ambiguous protonation states associated with the observed macroscopic pKas (SM07 and SM14). We discuss implications of the use of this experimental technique for the interpretation of pKa data, and provide suggestions for future pKa data collection efforts with the goal of evaluating or training computational pKa predictions.

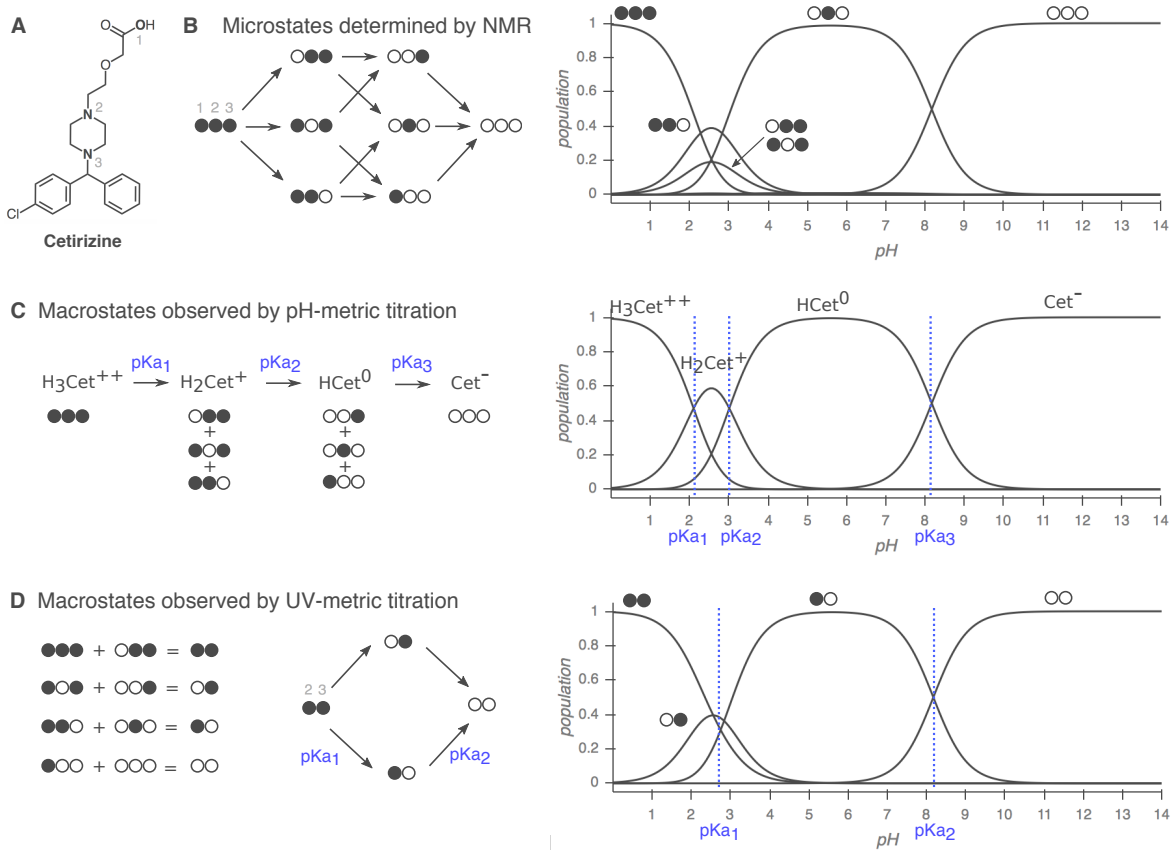

Figure 1. Comparison of macroscopic and microscopic pKa measurement methods. Filled circles represent protonated sites and empty circles represent deprotonated sites with the order of carboxylic acid (1), piperazine nitrogen (2), and piperazine nitrogen (3). Protonation state populations shown for pH-metric and UV-metric pKa measurement methods are simulations, calculated using NMR-based microscopic pKa values. (**A**) Cetirizine has n =3 titratable sites, shown in bold. (**B**) *Left:* The 8 microstates (2^n) and 12 micro-pKas ($n2^{n-1}$) of cetirizine. *Right:* Relative population of microspecies with respect to pH. Potentially all microstates can be resolved via NMR. (**C**) Simulated pH-metric (potentiometric) titration and macroscopic populations. For a polyprotic molecule, only macroscoopic pKas can be measured with pH-metric titration. Microstates that with different total charge (related to the number of protons) can be resolved, but microstates with the same total charge are observed as one macroscopic population. (**D**) Simulated microscopic populations for UV-metric (spectrophotometric) titration of cetirizine. Only protonation of the titration sites within four heavy atoms of the UV-chromophore is likely to cause an observable change in the UV-absorbance spectra, thus microstates that only differ by protonation of the distal carboxylic acid cannot be differentiated. Moreover, populations that overlap may or may not be resolvable depending on how much their absorbance spectra in the UV region differ. Both UV-metric and pH-metric pKa determination methods measure macroscopic pKas for polyprotic molecules, which cannot easily be assigned to individual titration sites and underlying microstate populations in the absence of other experimental evidence that provides structural resolution, such as NMR. Note that macroscopic populations observed in these two methods are composed of different combinations of microstates depending on the principles of measurement technique. Here, the illustrative diagram style was adopted from [10], and NMR-determined microscopic pKas for cetirizine were taken from [11].

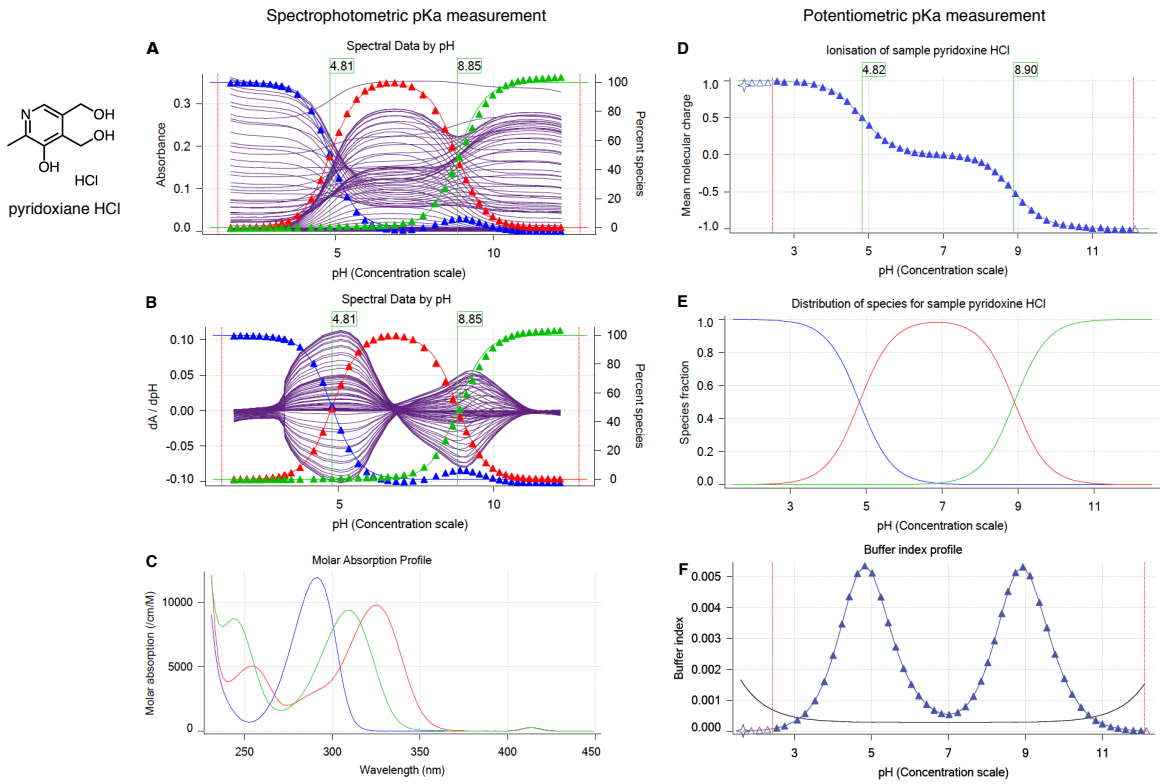

Figure 2. Spectrophotometric (UV-metric) and potentiometric (pH-metric) pKa measurements of pyridoxine HCl with Sirius T3. Spectrophotometic pKa measurement (panels **A, B, C**) relies on differences in the UV absorbance spectra between microscopic protonation states to deconvolute the population of macrostate species as a function of pH. While highly sensitive (and therefore requiring a very low analyte concentration of ~ 50 µM), this approach can only resolve changes in protonation states for titratible sites near chromophores and cannot separate the populations of microstates that change in the same manner as a function of pH. (**A**) Multiwavelength UV absorbance vs pH. Purple lines represents absorbance at distinct wavelengths in UV region. (**B**) Derivative of multiwavelength absorbance with respect to pH (dA/dpH) vs pH is plotted with purple lines. In **A** and **B** Blue, red, and green triangles represent population of protonation states (from most protonated to least protonated) as calculated from a global fit to experimental UV absorbances for all pH values, while thin lines denote model fits that utilize the fitted model pKas to compute populations. pKa values (green flags) correspond to inflection point of multiwavelength absorbance data where change in absorbance with respect to pH is maximum. (**C**) Molar absorption coefficients vs wavelength for each protonation state as resolved by TFA. **D, E, F** illustrate potentiometric pKa measurement where molar addition of acid or base is tracked as pH is titrated. (**D**) Mean molecular charge vs pH. Mean molecular charge is calculated based on the model provided for the analyte: predicted number and nature of titratable sites(acid or base type), and number of counter ions present. pKa values are calculated as inflection points of charge vs pH plot. (**E**) Predicted macroscopic protonation state populations vs pH calculated based on pKa values (H_2A^+: blue, HA: red, and A^-: green) (**F**) Buffering index vs pH profile of water (grey solid line, theoretical) and the sample solution (blue triangles represent experimental data points). A higher concentration of analyte (~ 5 mM) is necessary for the potentiometric method than the spectrophotometric method in order to provide large enough buffering capacity signal above water for an accurate measurement.

Methods

Compound selection and procurement

To select a set of small molecules focusing on the chemical space representative of kinase inhibitors for physicochemical property prediction challenges (pKa and lipophilicity) we started from the kinase-targeted subclass of the ZINC15 chemical library [19] and applied a series of filtering and selection rules as depicted in Figure 3A. We focused on the availability "now" and reactivity "anodyne" subsets of ZINC15 in the first filtering step [http://zinc15.docking.org/subclasses/kinase/substances/subsets/now+anodyne/]. The "now" label indicates the compounds were availabile for immediate delivery, while the "anodyne" label excludes compounds matching filters that flag compounds with the potential for reactivity or pan-assay interference (PAINs) [20, 21].

Next, we identified resulting molecules that were also available for procurement through eMolecules [22] (free version, downloaded 1 June 2017), the supplier that would be used for procurement in this exercise. To find the intersection of ZINC15 kinase subset and eMolecules database, we matched molecules using their canonical isomeric SMILES strings, as computed via the OpenEye OEChem Toolkit [23]. About 100 mg of each compound in powder form with at least 90% purity were purchased. We calculated 100 mg was enough for optimization and replicate experiments to measure pKa, logP and solubility measurements with Sirius T3. Each UV-metric and pH-metric pKa measurement requires minimum of 0.01 mg and 1.00 mg compound (solid or delivered via DMSO stock solution), respectively. logP and pH-dependent solubility measurements require around 2 mg and 10 mg of solid chemical.

To extract availability and price information from eMolecules, we queried using a list of SMILES (as reported in eMolecules database) of the intersection set. We further filtered the intersection set (1204 compounds) based on delivery time (Tier 1 suppliers, 2-week delivery) and at least 100 mg availability in powder form (format: Supplier Standard Vial).

Filtering for predicted measurable pKas and lack of experimental data

The Sirius T3 (Pion) instrument used to collect pKa and logP/logD measurements requires a titratable group in the pKa range of 2–12, so we aimed to select compounds with predicted pKas in the range of 3–11 to allow a ~1 pK unit margin of error in pKa predictions. pKa predictions for compound selection were calculated using Epik (Schödinger) sequential pKa prediction (scan) [24, 25] with target pH 7.0 and tautomerization allowed for generated states. We filtered out all compounds that did not have any pKas predicted between 3–11, as well as compounds with two predicted pKa values predicted to be less than 1 pKa unit apart to ensure individual pKas of multiprotic compounds could be resolved with spectrophotometric pKa measurements. With the goal of selecting compounds suitable for subsequent logP measurements, we eliminated compounds with OpenEye XlogP [26] values less than -1 or greater than 6. Subsets of compounds with molecular weights between 150–350 g/mol and 350–500 g/mol were selected for fragment-like and drug-like categories respectively. Compounds without available price or stock quantity information were eliminated. As the goal was to provide a blind challenge, compounds with publicly available experimental logP measurements were also removed. The sources we checked for experimental logP values were the following: DrugBank [27] (queried with eMolecules SMILES), ChemSpider [28] (queried by canonical isomeric SMILES), NCI Open Database August 2006 release [29], Enhanced NCI Database Browser [30] (queried with canonical isomeric SMILES), and PubChem [31] (queried with InChIKeys generated from canonical isomeric SMILES with NCI CACTUS Chemical Identifier Resolver [32].)

Filtering for kinase inhibitor-like scaffolds

In order to include common scaffolds found in kinase inhibitors, we analyzed the frequency of rings found in FDA-approved kinase inhibitors via Bemis-Murcko fragmentation [33, 34]. Heterocycles found more than once in FDA-approved kinase inhibitors are shown in Figure 3B. In selecting 25 compounds for the fragment-like set and 10 compounds for the drug-like set, we prioritized including at least one example of each heterocycle, although we failed to find compounds with piperazine and indazole that satisfied all other selection criteria. We observed that certain heterocycles (shown in Figure 3C) were overrepresented based on our selection criteria, thus limited the number of these structures in SAMPL6 challenge set to at most one

in each set. To achieve broad and uniform sampling of the measurable logP dynamic range, we segregated the molecules into bins of predicted XlogP values and selected compounds from each bin, prioritizing less expensive compounds.

Filtering for UV chromophores

The presence of UV chromophores (absorbing in 200–400 nm) in close proximity to protonation sites is necessary for spectrophotometric pKa measurements. To filter for molecules with UV chromophores, we looked at the substructure matches to the SMARTS pattern $[n,o,c][c,n,o]cc$ which were considered as the smallest unit of pi-conjugation that can constitute a UV chromophore. This SMARTS pattern describes 4 heavy atom extended conjugation systems composed of aromatic carbon, nitrogen, or oxygen, such as 1.3-butadiene, which possesses an absorption peak at 217 nm. Additionally, the final set of selected molecules was manually inspected to makes sure all potentially titratable groups were within 4 heavy atoms distance from a UV-chromophore.

25 fragment-like and 10 drug-like compounds were selected, out of which procurement of 28 were completed in time. pKa measurements for 17 (SM01-SM17) and 7 (SM18-SM24) were successful, respectively. The resulting set of 24 small molecules constitute the SAMPL6 pKa challenge set. The other 4 compounds, UV-metric pKa measurements did not detect pKas in the range of 2–12, so we decided not to include them in the SAMPL6 pKa challenge. Experiments for these 4 compounds are not reported in this publication.

Python scripts used in the compound selection process are available from GitHub [https://github.com/choderalab/sampl6-physicochemical-properties]. Procurement details for each compound can be found in Supplementary Table 1, and chemical properties used in selection of compounds are summarized in Supplementary Table 2.

UV-metric pKa measurements

Experimental pKa measurements were collected using the spectrophotometric pKa measurement method with a Sirius T3 automated titrator instrument (Pion) at 25°C and constant ionic strength. The Sirius T3 is equipped with an Ag/AgCl double-junction reference electrode to monitor pH, a dip probe attached to UV spectrophotometer, a stirrer, and automated volumetric titration capability. The Sirius T3 UV-metric pKa measurement protocol measures the change in multi-wavelength absorbance in the 250–450 nm UV region of the absorbance spectrum while the pH is titrated over pH 1.8–12.2 [12, 13]. Subsequent global data analysis identifies pH-dependent populations of macrostates and fits one or more pKa values to match this population with a pH-dependent model.

DMSO stock solutions of each compound with 10 mg/ml concentration were prepared by weighing 1 mg of powder chemical with a Sartorius Analytical Balance (Model: ME235P) and dissolving it in 100 µL DMSO (Dimethyl sulfoxide, Fisher Bioreagents, CAT: BP231-100, LOT: 116070, purity $\geq$ 99.7%). DMSO stock solutions were capped immediately to limit water absorption from the atmosphere due to the high hygroscopicity of DMSO, and sonicated for 5–10 minutes in a water bath sonicator at room temperature to ensure proper dissolution. These DMSO stock solutions were stored in room temperature up to 2 weeks in capped glass vials. 10 mg/ml DMSO solutions were used as stock solutions for the preparation of three replicate samples for independent titrations: For each experiment, 1–5 µL of 10 mg/ml DMSO stock solution was delivered to a 4 mL Sirius T3 glass sample vial with an electronic micropipette (Rainin EDP3 LTS 1-10 µL). The volume of delivered DMSO stock solution, which determines the sample concentration following dilution by the T3, is optimized individually for each compound to achieve sufficient but not saturated absorbance signal (targeting 0.5–1.0 AU) in the linear response region. Another limiting factor for sample concentration was ensuring that the compound remains soluble throughout the entire pH titration range. 25 µL of mid-range buffer (14.7 mM K_2HPO_4 0.15 M KCl in H_2O) was added to each sample, transfered with a micropipette (Rainin EDP3 LTS 10-100 µL). to provide enough buffering capacity in middle pH ranges so that pH could be controlled incrementally throughout the titration.

pH is temperature and ionic-strength dependent. A sample heating block on the Sirius T3 kept the analyte solution at 25 ± 0.5 °C throughout the titration. Sample ionic strength was adjusted by dilution in 1.5 mL ionic strength-adjusted water (ISA water, 0.15 M KCl) by the Sirius T3. Analyte dilution, mixing, acid/base titration,

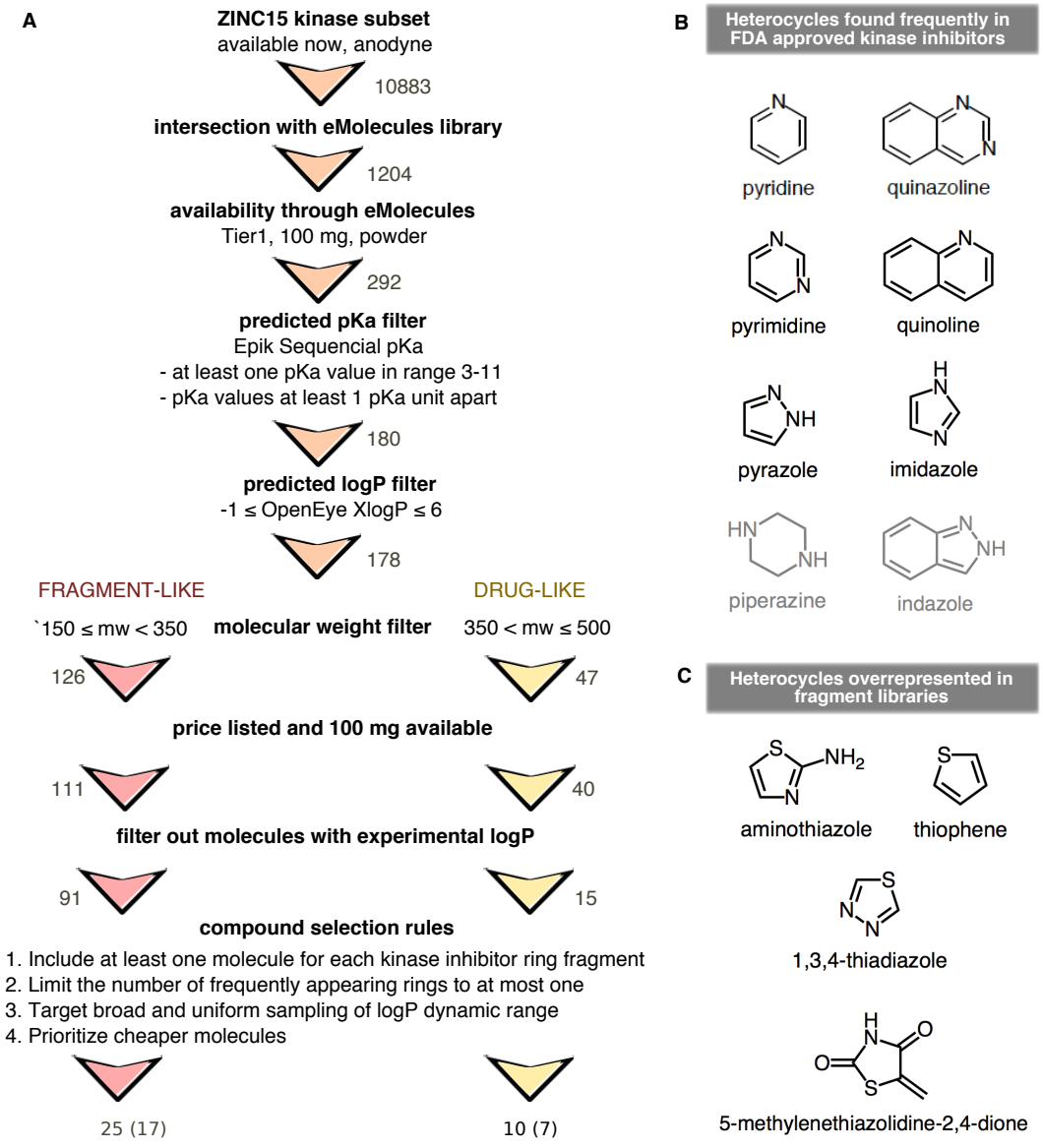

Figure 3. Compound selection for the SAMPL6 pKa challenge, with the goal of running subsequent logP/logD challenges on the same compound set. (**A**) Flowchart of filtering steps for the selection of compounds that resemble kinase inhibitors and their fragments. Numbers next to arrows indicate the number of compounds remaining after each filtering step. A total of 25 fragment-like and 10 drug-like compounds were selected, out of which procurement and pKa measurements for 17 and 7 were successful, respectively. (**B**) Frequent heterocycles found in FDA approved kinase inhibitors, as determined by Bemis-Murcko fragmentation into rings [33]. Black structures were represented in SAMPL6 set at least once. Compounds with piperazine and indazole (gray structures) could not be included in the challenge set due to library and selection limitations. (**C**) Structures of heterocycles that were overrepresented based on our compound selection workflow. We have limited occurance of these heterocycles to at most one.

and measurement of UV-absorbance was automated by the Sirius T3 UV-metric pKa measurement protocol. The pH was titrated between pH 1.8 and 12.2 via the addition of acid (0.5 M HCl) and base (0.5 M KOH), targeting 0.2 pH steps between UV absorbance spectrum measurements. Titrations were performed under argon flow on the surface of the sample solution to limit the absorption of carbon dioxide from air, which can alter the sample pH to a measurable degree. To fully capture all sources of experimental variability, instead of performing three sequential pH titrations in the same sample solution, three replicate samples (prepared from the same DMSO stock solution) were subjected to one round of pH titration each. Although this choice reduced throughput and increased analyte consumption, it limited the dilution of the analyte during multiple titrations, resulting in stronger absorbance signal for pKa fitting. Under circumstances where analyte is scarce, it is also possible to do three sequential titrations using the same sample to limit consumption when the loss of accuracy is acceptable.

Visual inspection of the sample solutions after titration and inspection of the pH-dependent absorbance shift in the 500–600 nm region of the UV spectra was used to verify no detectable precipitation occurred during the course of the measurement. Increased absorbance in the 500–600 nm region of the UV spectra is associated with scattering of longer wavelengths of light in the presence of colloidal aggregates. For each analyte, we optimized analyte concentration, direction of the titration, and pH titration range in order to maintain solubility over the entire experiment. The titration direction was specified so that each titration would start from the pH where the compound is most soluble: low-to-high pH for bases and high-to-low pH for acids. While UV-metric pKa measurements can be performed with analyte concentrations as low as 50 µM (although this depends on the absorbance properties of the analyte), some compounds may yet not be soluble at these low concentrations throughout the pH range of the titration. As the sample is titrated through a wide range of pH values, it is likely that low-solubility ionization states—such as neutral and zwitterionic states—will also be populated, limiting the highest analyte concentration that can be titrated without encountering solubility issues. For compounds with insufficient solubility to accurately determine a pKa directly in a UV-metric titration, a cosolvent protocol was used, as described in the next section (**UV-metric pKa measurement with cosolvent**).

Two Sirius T3 computer programs—Sirius T3 Control v1.1.3.0 and Sirius T3 Refine v1.1.3.0—were used to execute measurement protocols and analyze pH-dependent multivelenghth spectra, respectively. Protonation state changes at titratable sites near chromophores will modulate the UV-absorbance spectra of these chromophores, allowing populations of distinct UV-active species to be resolved as a function of pH. To do this, basis spectra are identified and populations extracted via TFA analysis of the pH-dependent multi-wavelength absorbance [13]. When fitting the absorbance data to a titratable molecule model to estimate pKas, we selected the minimum number of pKas sufficient to provide a high-quality fit between experimental and modeled data based on visual inspection of pH-dependent populations.

This method is capable of measuring pKas between 2–12 when protonatable groups are at most 4–5 heavy atoms away from chromophores such that a change in protonation state alters the absorbance spectrum of the chromophore. We selected compounds where titratable groups are close to potential chromophores (generally aromatic ring systems), but the possibility exists that our experiments did not detect protonation state changes of titratable groups distal from UV chromophores.

UV-metric pKa measurement with cosolvent

If analytes are not sufficiently soluble in aqueous environments, pKa values cannot be accurately determined with UV-metric pKa measurements. If precipitation occurs, the UV-absobance signal from pH-dependent precipitate formation can not be differentiated from the pH-dependent signal of soluble microstate species. For compounds with low aqueous solubility, pKa values were estimated from multiple apparent pKas measurements performed in methanol:water cosolvent solutions with various mole fractions, from which the pKa at 0% methanol (pure water) can be extrapolated. This method is referred to as a UV-metric psKa measurement in the Sirius T3 Manual [35].

The cosolvent spectrophotometric pKa measurement protocol was very similar to the standard aqueous UV-metric pKa measurement protocol, with the following differences: Titrations were performed in typically in 30%, 40%, and 50% mixtures methanol:water by volume to measure apparent pKa values (psKa) in these

mixtures. Yasuda-Shedlovsky extrapolation was subsequently used to estimate the pKa value at 0% cosolvent (Figure 4) [15, 36, 37].

$$\mathrm{psKa} + \log[\mathrm{H_2O}] = A/\epsilon + B \tag{1}$$

Yasuda-Shedlovsky extrapolation relies on the linear correlation between $\mathrm{psKa} + \log[\mathrm{H_2O}]$ and the reciprocal dielectric constant of the cosolvent mixture ($1/\epsilon$). In Eq. 1, A and B are the slope and intercept of the line fitted to experimental datapoints. Depending on the solubility requirements of the analyte, the methanol ratio of the cosolvent mixtures was adjusted. We designed the experiments to have at least 5% cosolvent ratio difference between datapoints and no more than 60% methanol content. Calculation of the Yasuda-Shedlovsky extrapolation was performed by the Sirius T3 software using at least 3 psKa values measured in different ratios of methanol:water. Addition of methanol (ISA, 0.15 M KCl) was controlled by the instrument before each titration. Three consecutive pH titrations at different methanol concentrations were performed using the same sample solution. In addition, three replicate measurements with independent samples (prepared from the same DMSO stock) were collected.

Calculation of uncertainty in pKa measurements

Experimental uncertainties were reported as the standard error of the mean (SEM) of three replicate pKa measurements. Standard deviation(σ) is calculated using Equation 2 using 0 as Delta Degrees of Freedom (ddof) value, to calculate maximum likelihood estimate of the variance of normally distributed variables. x_i are observations, $\overline{x}$ is the mean of observations and N is the number of observations. SEM is calculated as shown in Equation 3.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N - ddof}} \tag{2}$$

$$SEM = \frac{\sigma}{\sqrt{N}} \tag{3}$$

Since the Sirius T3 software reports pKa values to only two decimal places, we have reported SEM as 0.01 in cases where SEM values calculated from 3 replicates were lower than 0.01. SEM calculated from replicate measurements were found to be larger than non-linear fit error reported by the Sirius T3 Refine Software from UV-absorbance model fit of a single experiment, thus leading us to believe that running replicate measurements and reporting mean and SEM of pKa measurements is better for capturing all sources of experimental uncertainty. Notably, for UV-metric measurements, the measured pKa values should be insensitive to final analyte concentration and any uncertainty in the exact analyte concentration of the original DMSO stock solution, justifying the use of the same stock solution (rather than independently prepared stock solutions) for multiple replicates.

Protonation site determination with NMR

TO DO: NMR method.

Quality control for chemicals

Compound purity was assessed by LC-MS using an Agilent HPLC1200 Series equipped with auto-sampler, UV diode array detector, and a Quadrupole MS detector 6140. Chemistation version C01.07SR2 was used to analyze LC & LC/MS. A Assentis Express C18 column with 3.0x100 mm 2.7 µl particle size was used, with column temperature = 45°C.

- Mobile phase A: 2 mM ammonium formate (pH=3.5) aqueous
- Mobile phase B: 2 mM ammonium formate in Acetonitrile : Water=90:10 (pH=3.5)
- Flow rate : 0.75 ml/min
- Gradient: Starting with 10% B to 95%B in 10 minutes then hold 95%B for 5 minutes.
- Post run length: 5 minutes
- Mass condition: ESI positive and negative mode

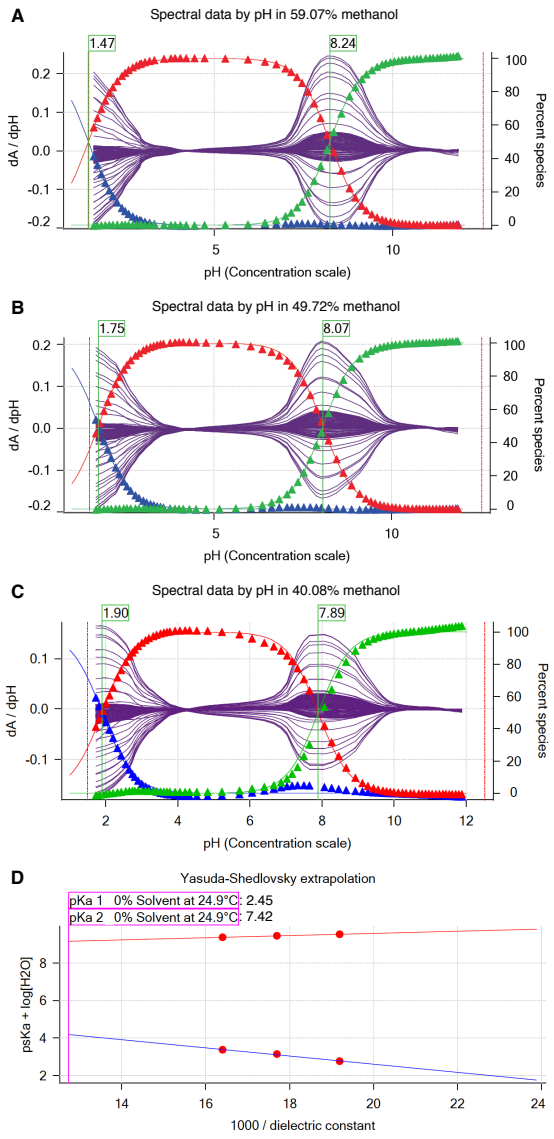

Figure 4. Determination of SM22 pKa values with cosolvent method and Yasuda-Shedlovsky extrapolation. A, **B**, and **C** show psKa of SM22 determined at various methanol concentrations: 59.07%, 49.72%, 40.08% by weight. Purple solid lines indicate derivative of absorbance with respect to pH vs pH in multiple wavelength. psKa values (green flags) were determined by Sirius T3 Refine Software. Blue, red and green triangles show relative populations of macroscopic protonation states with respect to pH calculated from the experimental data. Notice that as cosolvent concentration increases, psKa1 value decreases from 1.90 to 1.47 and psKa2 value increases from 7.84 to 8.24. **D** Yasuda-Shedlovsky extrapolation plot for SM22. Red datapoints correspond to psKa determined at various cosolvent ratios. Based on linear fitting to $psKa + log[H_2O]$ vs $1/\epsilon$, pKa1 and pKa2 in 0% cosolvent (aqueous solution) was determined as 2.45 and 7.42, respectively. R^2 values of linear fits are both 0.99. The slope of Yasuda-Shedlovsky extrapolation shows if the observed titration has acidic(positive slope) or basic(negative) character dominantly, although this is an macroscopic observation and should not be relied on for annotation of pKas to functional groups (microscopic pKas).

- Capillary voltage: 3000 V
- Grying gas flow: 12 ml/min
- Nebulizer pressure: 35 psi
- Drying temperature: 350°C
- Mass range: 5-1350 Da; Fragmentor:70; Threshold:100

The percent area for primary peak is calculated based on the area of the peak divided by the total area of all peaks. The percent area of the primary peak is reported as an estimate of sample purity.

Results

Spectrophotometric pKa measurements

Spectrophotometrically-determined pKa values for all molecules from the SAMPL6 pKa challenge are shown in Figure 5 and Table 1. The protocol used—cosolvent or aqueous UV-metric titration—is indicated in Table 1 together with SEM of each reported measurement. Out of 24 molecules successfully assayed, five molecules have two resolvable pKa values, while one has three resolvable pKa values within the measurable range 2–12 pKa range. The SEM of reported pKa measurements is low, with the largest uncertainty reported being 0.04 pK units (pKa1 of SM06 and pKa3 of SM18). Individual replicate measurements can be found in Supplementary Table 3. Reports generated for each pKa measurement by the Sirius T3 Refine software can also be found in the Supplementary Information. Experimental pKa values for nearly all compounds with multiple resolvable pKas are well-separated (more than 3 pKa units), except for SM14 and SM18.

Impact of cosolvent to UV-metric pKa measurements

For molecules with insufficient aqueous solubilities throughout the titration range (pH 2-12), we resorted to cosolvent UV-metric pKa measurements, with methanol used as cosolvent. To confirm that cosolvent UV-metric pKa measurements led to indistinguishable results compared to aqueous UV-metric measurements, we collected pKa values of 12 highly soluble SAMPL6 compounds—as well as pyridoxine—-using both cosolvent and aqueous methods. Correlation analysis of pKa values determined with both methods demonstrated that using methanol as cosolvent and determining aqueous pKas via Yasuda-Shedlovsky extrapolation did result in significant bias (Figure 6). Means and standard errors of UV-metric pKa measurements with and without cosolvent are provided in Supplementary Table 5. pKa measurement results of individual replicate measurements with and without cosolvent can be found in Supplementary Table 4.

Purity of SAMPL6 compounds

LC-MS based purity measurements showed that powder stocks of 23 of the SAMPL6 pKa challenge compounds were >90% pure, while purity of SM22 was 87%—the lowest in the set (Supplementary Table 6).

JDC: Can we say something about how accurate the LC-MS purity determination measurements are expected to be? 1%? Better than that?

Additionally, molecular weights detected by LC-MS method were consistent with those reported in eMolecules, as well as supplier-reported molecular weights, when provided. The Sirius T3 protocols recommend compound purities ~90% purity to minimize the impact on high-accuracy pKa measurements.

Characterization of SM07 microstates with NMR

Add results of NMR analysis of SM07

Characterization of SM14 microstates with NMR

Add results of NMR analysis of SM07

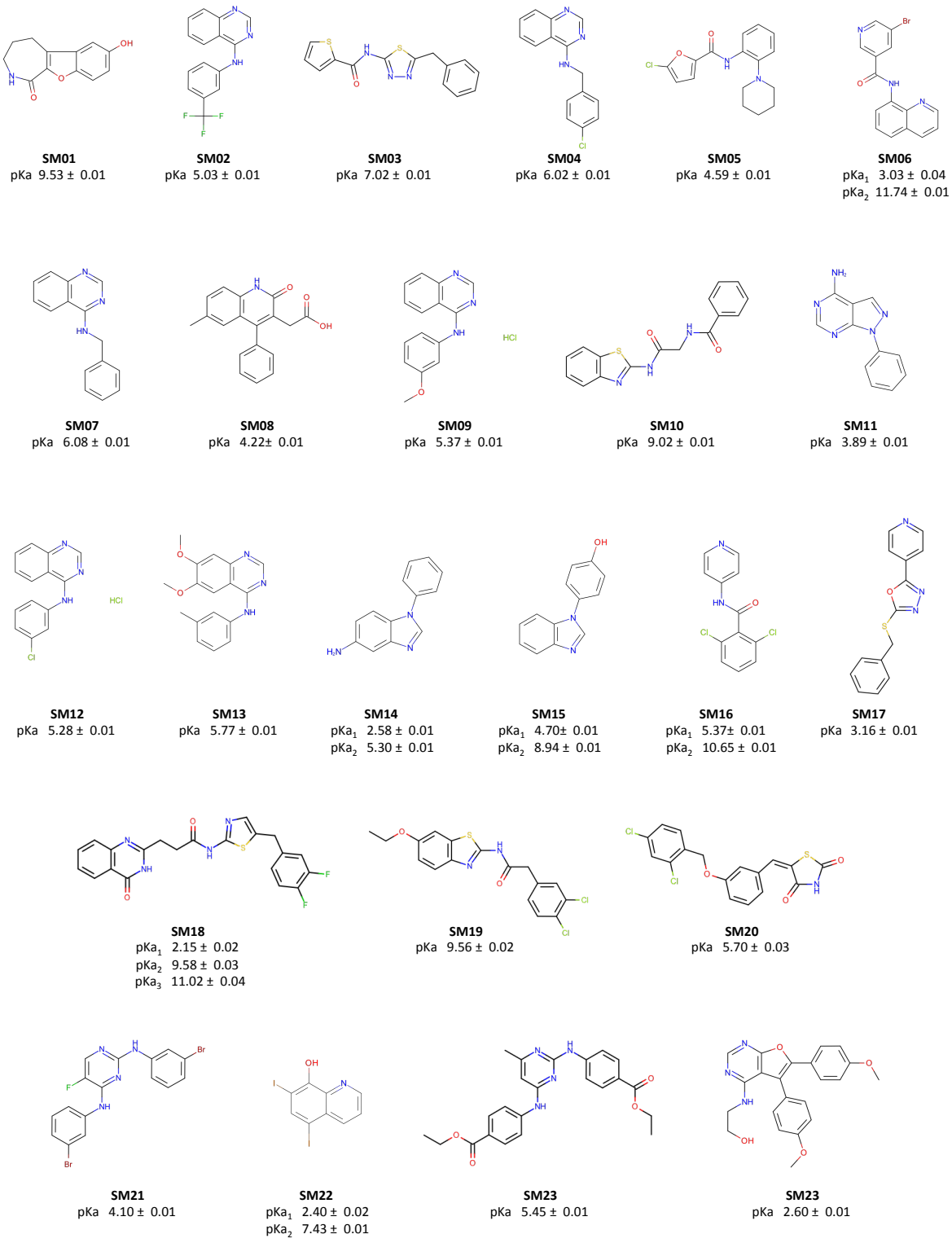

Figure 5. Molecules used in the SAMPL6 pKa challenge. Experimental UV-metric pKa measurements were performed for these 24 molecules, and discernable macroscopic pKas reported. Uncertainties are expressed as the standard error of the mean (SEM) of three independent measurements. 2D structures were created with OpenEye OEDepict Toolkit [38]

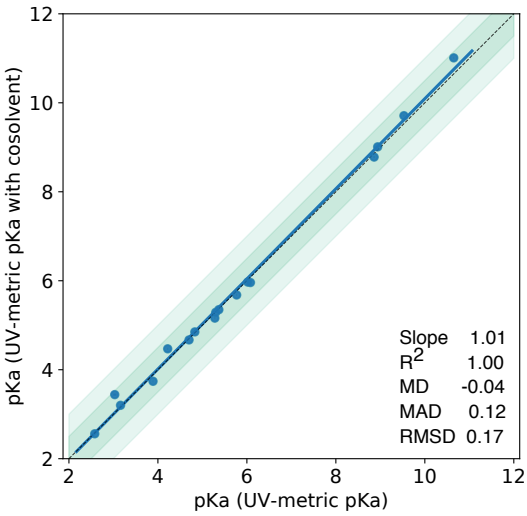

Figure 6. pKa measurements with UV-metric method with cosolvent and UV-metric method in water show good correlation. 17 pKa values(blue marks) of 13 chemicals were measured with both UV-metric pKa method in water and UV-metric pKa method with methanol as cosolvent (Yasuda-Shedlovsky extrapolation to 0% methanol). Dashed black line has slope of 1, representing perfect correlation. Dark and light green shaded areas indicate +-0.5 and +-1.0 pKa unit difference regions, respectively. Error bars are plotted as SEM of replicate measurements, although they are not visible since the largest SEM value is 0.04. MD: Mean difference, MAD: Mean absolute deviation, RMSD: Root-mean-square deviation.

JDC: Can we compute bootstrapped error bars for these statistics?

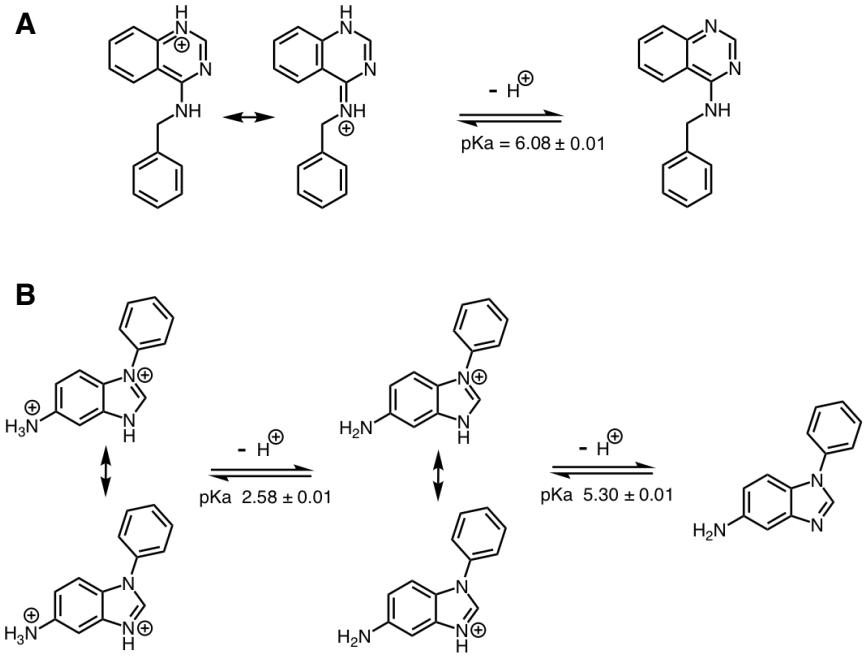

Figure 7. Microstates of SM07 and SM14 characterized by NMR. pKa values were determined by UV-metric titrations. NMR characterization showed that experimental pKas observed with spectrophotometric method were related microscopic transitions between illustrated microstates of **A** SM07 and **B** SM14.

Table 1. Experimental pKas of SAMPL6 compounds. Spectrophotometric pKa measurements were performed with two assay types based on aqueous solubility of analytes. "UV-metric pKa" assay indicates spectrophotometric pKa measurements done with Sirius T3 in ISA water. "UV-metric pKa with cosolvent" assay refers to pKa determination by Yasuda-Shedlovsky extrapolation from psKa measurements in various ratios of ISA methanol:water mixtures. Triplicate measurements were performed at 25 ± 0.5 °C and in the presence of approximately 150 mM KCl to adjust ionic strength.

Molecule ID	pKa_1	pKa_2	pKa_3	Assay Type
SM01	9.53 ± 0.01			UV-metric pKa
SM02	5.03 ± 0.01			UV-metric pKa with cosolvent
SM03	7.02 ± 0.01			UV-metric pKa with cosolvent
SM04	6.02 ± 0.01			UV-metric pKa
SM05	4.59 ± 0.01			UV-metric pKa with cosolvent
SM06	3.03 ± 0.04	11.74 ± 0.01		UV-metric pKa
SM07	6.08 ± 0.01			UV-metric pKa
SM08	4.22 ± 0.01			UV-metric pKa
SM09	5.37 ± 0.01			UV-metric pKa with cosolvent
SM10	9.02 ± 0.01			UV-metric pKa with cosolvent
SM11	3.89 ± 0.01			UV-metric pKa
SM12	5.28 ± 0.01			UV-metric pKa
SM13	5.77 ± 0.01			UV-metric pKa
SM14	2.58 ± 0.01	5.30 ± 0.01		UV-metric pKa
SM15	4.70 ± 0.01	8.94 ± 0.01		UV-metric pKa
SM16	5.37 ± 0.01	10.65 ± 0.01		UV-metric pKa
SM17	3.16 ± 0.01			UV-metric pKa
SM18	2.15 ± 0.02	9.58 ± 0.03	11.02 ± 0.04	UV-metric pKa with cosolvent
SM19	9.56 ± 0.02			UV-metric pKa with cosolvent
SM20	5.70 ± 0.03			UV-metric pKa with cosolvent
SM21	4.10 ± 0.01			UV-metric pKa with cosolvent
SM22	2.40 ± 0.02	7.43 ± 0.01		UV-metric pKa with cosolvent
SM23	5.45 ± 0.01			UV-metric pKa with cosolvent
SM24	2.60 ± 0.01			UV-metric pKa with cosolvent

[1] pKa values are reported as mean ± SEM of three replicates.

Discussion

Sample preparation and effect of cosolvents

Samples for UV-metric pKa measurements were prepared by dilution of up to 5 µL DMSO stock solution of analyte in 1.5 mL ISA water, which results in the presence of ~0.3% DMSO during titration, which has negligible effect to pKa measurements. For UV-metric or pH-metric measurements, it is possible to prepare samples without DMSO, but it is difficult to prepare samples by weighing extremely low amounts of solid stocks (in the order of 0.01–0.10 mg) to target 50 µM analyte concentrations, even with an analytical balance. For experimental throughput, we therefore preferred using DMSO stock solutions. Another advantage of starting from DMSO stock solutions is that it helps to overcome kinetic solubility problems of analytes.

In UV-metric measurements, both water and methanol (when used as cosolvent) stock solutions were ionic strength adjusted with 150 mM KCl, but acid and base solutions were not. This means that throughout pH titration ionic strength slighly fluctuates, but on average ionic strength of samples were staying around 150–180 mM. By using ISA solutions the effect of salt concentration change on pKa measurements were minimized.

If an analyte is soluble enough, UV-metric pKa measurements in water should be prefered over cosolvent methods, since pKa measurement in water is more direct. For pKa determination via cosolvent extrapolation

using methanol, apparent pKas in at least three different methanol:water ratios must be measured, and pKa in 0% cosolvent computed by Yasuda-Shedlovsky extrapolation. The number and spread of psKa measurements and error in linear fit extrapolation influences the accuracy of pKas determined by this approach. To test that UV-metric methods with or without cosolvent have indistinguishable performance, we collected pKa values for 17 SAMPL6 compounds and pyridoxine with both methods. Figure 6 shows there is good correlation between both methods, and that the mean absolute deviation between two methods is 0.12. Mean deviation between the two sets is -0.04, showing there is no bias in cosolvent measurements.

TO DO: Add bootstrapped uncertainty or CI to mean deviation value.

Impact of impurities

Precisely how much the presence of small amounts of impurities impacts UV-metric pKa measurements is unpredictable. For an impurity to alter UV-metric pKa measurements, it must possess a UV-chromophore and a titratable group in the vicinity of the chromophore—otherwise, it would not interfere with absorbance signal of the analyte. If a titratable impurity *does* possess a UV-chromophore, UV multiwavelenghth absorbance from the analyte and impurity will be convoluted. The impact the impurity will have on the multiwavelength absorbance spectra depends on the strength of its molar absorption coefficient. In the worst case schenario, an impurity with high concentration or strong UV absorbance can shift the measured pKa value or create the appearance of an extra pKa. As a result, it is important to use analytes with high purities to obtain high accuracy pKa measurements. Therefore, we confirmed the purities of SAMPL6 compounds with LC-MS.

Interpretation of UV-metric pKa measurements

Multiwavelength absorbance analysis on the Sirius T3 allows for very good resolution of pKas, but it is important to note that this method produces *macroscopic* pKas for polyprotic compounds. If multiple microscopic pKas have close pKa values and overlapping changes in UV absorbance spectra associated with protonation/deprotonation, the spectral analysis could produce a single macroscopic pKa that represents an aggregation of multiple microscopic pKas. An extreme example of such case is demonstrated in the simulated macrostate populations of cetirizine that would be observed with UV-metric titration (Figure 1).

If protonation state populations observed via UV-metric titrations (such as in Figure 2B) are composed of single microstate, experimentally measured pKas are indeed microscopic pKas. Unfortunately, judging the composition of experimental populations is not possible by just using UV-metric or pH-metric titration. Molecules in the SAMPL6 pKa challenge dataset with only one pKa measured in the 2–12 range could therefore be monoprotic (possessing a single titratable group that changes protonation state by gain or loss of a single proton over this pH range) or polyprotic (gaining or losing multiple protons from one or more sites with overlapping microscopic pKa values). Similarly, titration curves of molecules with multiple experimental pKas may show well-separated microscopic pKas or macroscopic experimental pKas that are really composites of microscopic pKas with similar values. Therefore, without additional experimental evidence, UV-metric pKas should not be assigned to individual titratable groups.

Determination of the exact microstates populated at different pH values via NMR can provide a complementary means of differentiating between microscopic and macroscopic pKas in cases where there is ambiguity. As determination of protonation microstates via NMR is very laborious, we were only able to characterize microstates of two molecules: SM07 and SM11.

In UV-metric pKa measurements with cosolvent, the slope of Yasuda-Shedlowsky extrapolation can interpreted to understand if pKa has dominantly acidic or basic character. As the methanol ratio is increased, psKa values of acids increase, while psKa values for bases decrease. However, it is important to remember that if measured pKa is macroscopic, acid/base assignment from cosolvent psKa trends is also a macroscopic property, and should not be used as a guide for assigning pKa values to functional groups [39].

NMR microstate characterization

TO DO: Add discussion of NMR results

The goal of NMR characterization was to collect information on microscopic states related to experimental pKa measurements, i.e., determine exact sites of protonation. pKa measurements performed with spectrophotometric method provide macroscopic pKa values, but do not provide information on the site(s) of protonation. On the other hand, most computational prediction methods primarily predict microscopic pKa values. Protonation sites can be determined by NMR methods, although these measurements are very laborious in terms of data collection and interpretation compared to pKa measurements with the autoamted Sirius T3. Moreover, not all SAMPL6 molecules were suitable for NMR measurements due to the high sample concentration requirements (for methods other than proton NMR) and analyte solubility issues. We performed NMR based microstate characterization only for SM07 and SM14.

We investigated microstates populated at pH values lower and higher than macroscopic pKa value, with the goal of evaluating if UV-metric pKa was microscopic (related to single protonation site).

JDC: Can you really see microstate populations, or just proton average populations? If you had two tautomer species where the proton flips back and forth between two sites but were populated 50%-50%, would that be indistinguishable from one species populated at 50% that had both protons?

In addition to SM07, there were five other 4-amino quinazoline derivatives in SAMPL6 set: SM02, SM04. SM09, SM12, and SM13. Based on structural similarity, we can infer that spectrophotometric pKa values measured for other 4-amino quinazoline compounds as microscopic pKa related to the protonation of the same quinazoline nitrogen with the same neutral background protonation states.

TO DO: Any additional challenges worth mentioning with NMR measurments?

Heavy atom spectra that rely on natural isotope abundance requires high sample concentrations (preferably in the order of 100 mM). It is possible that drug or drug-fragment-like compounds, such as the compounds used in this study, have insufficient aqueous solubility, limiting the choice of solvent and pH. It may be necessary to use organic cosolvents to prepare these high concentration solutions, or only prepare samples at pH values that correspond to high solubility states (e.g., when charged state of the compound is populated).

Suggestions for future pKa data collection for computational challenges

Most high throughput pKa measurement methods measure macroscopic pKas. One way to circumvent this problem is to confine our interest in future pKa challenges to experimental datasets containing only monoprotic compounds if UV-metric or potentiometric pKa measurements can be made. Then it would be possible to unambiguously assign pKa values to underlying protonation states. However, it is important to consider that multiprotic compounds are common in pharmaceutically interesting molecules, and it is important to be able to model them reliably.

Although relatively efficient, UV-metric pKa measurements with the Sirius T3 do not provide structural information about microstates. Even the acid-base assignment based on direction of psKa shift with cosolvent is not a reliable indicator for assigning experimental pKas to individual functional groups in in multiprotic compounds. On the other hand, most computational pKa prediction methods output microscopic pKas. It is therefore difficult to use experimental macroscopic pKa values to assess and train microscopic pKa prediction methods directly without further means of annotating macroscopic-microscopic correspondence. It is usually impossible to infer the underlying microscopic pKa values from macroscopic measurements of a polyprotic compound without complementary experiments that can provide structural information.

Another source of complexity is how the observed macroscopic pKas, i.e., the situations under which microstates become indistinguishable, changes between different experimental methods as illustrated in Figure 1. The "macroscopic" label is commonly ascribed to transitions between different ionization states of a molecule (all microstates that have the same total charge form one macrostate), but this definition only applies to potentiometric methods. In UV-absorbance based methods, the principle that determines which microstates will be distinguishable is not charge or number of bound protons, but molecular absorbance changes, and how closely undelying microscopic pKa values overlap. To compare experimental macroscopic pKa and microscopic computational predictions on common ground, the best solution is to compute "predicted" macroscopic pKa values from microscopic pKas following the principle of experiment type. A

disadvantage of this approach is that experimental data cannot provide direct guidance on microscopic pKa resolution for improving pKa prediction methods.

Since analyte purity is critical for accuracy, necessary quality control experiments must be performed to ensure at least 90% purity for UV-metric pKa measurements. Higher purity values may be necessary for other methods. For potentiometric methods, knowing the stoichiometry of any counterions present in the original powder stocks is also necessary.

For the set of SAMPL6 pKa challenge compounds, we could not use potentiometric pKa measurements due to the low aqueous solubility of many of these compounds. The lowest solubility observed *somewhere* in the experimental pH range of titration is the limiting factor, since for accurate measurements the analyte must stay in the solution phase throughout the whole titration. Since the titration pH range is determined with the goal of capturing all ionization states, the analyte is inevitably exposed to pH values that correspond to low solubility. Neutral and zwitterionic species can be orders of magnitude less soluble than ionic species.

For future pKa challenges with multiprotic compounds, if sufficient time and effort can be spared, it would be ideal to construct an experimental pKa dataset using experimental methods that can measure microscopic pKas directly, such as NMR. In this study, we were only able to perform follow up NMR microstate characterization of two compounds due to the labor-intensive nature of NMR data collection and analysis.

In the future pKa challenges, it would be especially interesting to expand this exercise to larger and more flexible drug-like molecules. pKa values are environment dependent and it would be useful to be able to predict pKa shifts based on on ionic strength, temperature, lipophilic content, with cosolvents or in organic solvents. Measuring the pKa of molecules in organic solvents would be useful for guiding process chemistry. To test such predictions, special pKa experiments would need to be designed to measure pKas under different conditions.

Code and data availability

- SAMPL6 pKa challenge instructions, submissions, experimental data and analysis is available at https://github.com/MobleyLab/SAMPL6
- Python scripts used for compound selection are available at **compound_selection** directory of https://github.com/choderalab/sampl6-physicochemical-properties

Construct a proper README for compound selection directory.

Overview of supplementary information

Organized in SI document:

- TABLE SI 1: procurement details of SAMPL6 compounds
- TABLE SI 2: selection details of SAMPL6 compounds
- TABLE SI 3: pKa results of replicate experiments csv
- TABLE SI 4: pKa results of water and cosolvent replicate experiments csv
- TABLE SI 5: pKa mean and SEM results of water and cosolvent replicate experiments
- NMR spectra of SM07 microstate characterization
- TABLE SI 6: Summary of LC-MS purity results
- LC-MS Figures

Extra files:

- Sirius T3 reports

Author Contributions

Conceptualization, MI, JDC, TR, ASR, DM ; Methodology, MI, DL, IEN ; Software, MI, ASR ; Formal Analysis, MI ; Investigation, MI, DL, IEN, HW, XW, MR; Resources, TR, DL; Data Curation, MI ; Writing-Original Draft, MI, JDC, IEN; Writing - Review and Editing, MI, DL, ASR, IEN, HW, XW, MR, GEM, DM, TR, JDC; Visualization, MI, IEN ; Supervision, JDC, TR, DM, GEM ; Project Administration, MI ; Funding Acquisition, JDC, DM, TR, MI.

Acknowledgments

Complete this section.

MI, ASR, and JDC acknowledge support from the Sloan Kettering Institute. JDC acknowledges support from NIH grant P30 CA008748. MI acknowledges Doris J. Hutchinson Fellowship. The authors are extremely grateful for the assistance and support from the Merck Preformulations and NMR Structure Elucidation groups for materials, expertise, and instrument time, without which this SAMPL challenge would not have been possible. MI, ASR, and JDC are grateful to OpenEye Scientific for providing a free academic software license for use in this work. MI is grateful to Pion/Sirius Analytical for their technical support in the planning and execution of this study.

Brad Sherborne

Paul Czodrowski - helped with challence construction and purchasable compound list in earlier iteration

Caitlin Bannan - feedback on experimental data collection from computational perspective

Disclosures

JDC is a member of the Scientific Advisory Board for Schrödinger, LLC.

References

[1] **Mobley DL**, Chodera JD, Isaacs L, Gibb BC. Advancing predictive modeling through focused development of model systems to drive new modeling innovations. UC Irvine: Department of Pharmaceutical Sciences, UCI. 2016; https://escholarship.org/uc/item/7cf8c6cr.

[2] **Drug Design Data Resource**, SAMPL;. https://drugdesigndata.org/about/sampl.

[3] **Bannan CC**, Burley KH, Chiu M, Shirts MR, Gilson MK, Mobley DL. Blind prediction of cyclohexane–water distribution coefficients from the SAMPL5 challenge. Journal of Computer-Aided Molecular Design. 2016 Nov; 30(11):927–944. http://link.springer.com/10.1007/s10822-016-9954-8, doi: 10.1007/s10822-016-9954-8.

[4] **Rustenburg AS**, Dancer J, Lin B, Feng JA, Ortwine DF, Mobley DL, Chodera JD. Measuring experimental cyclohexane-water distribution coefficients for the SAMPL5 challenge. Journal of Computer-Aided Molecular Design. 2016 Nov; 30(11):945–958. http://link.springer.com/10.1007/s10822-016-9971-7, doi: 10.1007/s10822-016-9971-7.

[5] **Pickard FC**, König G, Tofoleanu F, Lee J, Simmonett AC, Shao Y, Ponder JW, Brooks BR. Blind prediction of distribution in the SAMPL5 challenge with QM based protomer and pK a corrections. Journal of Computer-Aided Molecular Design. 2016 Nov; 30(11):1087–1100. http://link.springer.com/10.1007/s10822-016-9955-7, doi: 10.1007/s10822-016-9955-7.

[6] **Darvey IG**. The assignment of pKa values to functional groups in amino acids. Wiley Online Library; 1995.

[7] **Bezençon J**, Wittwer MB, Cutting B, Smieško M, Wagner B, Kansy M, Ernst B. pKa determination by 1H NMR spectroscopy – An old methodology revisited. Journal of Pharmaceutical and Biomedical Analysis. 2014 May; 93:147–155. http://linkinghub.elsevier.com/retrieve/pii/S0731708513005992, doi: 10.1016/j.jpba.2013.12.014.

[8] **Elson EL**, Edsall JT. Raman spectra and sulfhydryl ionization constants of thioglycolic acid and cysteine. Biochemistry. 1962; 1(1):1–7.

[9] **Elbagerma MA**, Edwards HGM, Azimi G, Scowen IJ. Raman spectroscopic determination of the acidity constants of salicylaldoxime in aqueous solution. Journal of Raman Spectroscopy. 2011 Mar; 42(3):505–511. http://doi.wiley.com/10.1002/jrs.2716, doi: 10.1002/jrs.2716.

[10] **Rupp M**, Korner R, V Tetko I. Predicting the pKa of small molecules. Combinatorial chemistry & high throughput screening. 2011; 14(5):307–327.

[11] **Marosi A**, Kovács Z, Béni S, Kökösi J, Noszál B. Triprotic acid–base microequilibria and pharmacokinetic sequelae of cetirizine. European Journal of Pharmaceutical Sciences. 2009 Jun; 37(3-4):321–328. http://linkinghub.elsevier.com/retrieve/pii/S0928098709000773, doi: 10.1016/j.ejps.2009.03.001.

[12] **Tam KY**, Takács-Novák K. Multi-wavelength spectrophotometric determination of acid dissociation constants: a validation study. Analytica chimica acta. 2001; 434(1):157–167.

[13] **Allen RI**, Box KJ, Comer JEA, Peake C, Tam KY. Multiwavelength spectrophotometric determination of acid dissociation constants of ionizable drugs. Journal of pharmaceutical and biomedical analysis. 1998; 17(4):699–712.

[14] **Comer JEA**, Manallack D. Ionization Constants and Ionization Profiles. In: *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* Elsevier; 2014.http://linkinghub.elsevier.com/retrieve/pii/B9780124095472112338, doi: 10.1016/B978-0-12-409547-2.11233-8.

[15] **Avdeef A**, Box KJ, Comer JEA, Gilges M, Hadley M, Hibbert C, Patterson W, Tam KY. PH-metric logP 11. pK a determination of water-insoluble drugs in organic solvent–water mixtures. Journal of pharmaceutical and biomedical analysis. 1999; 20(4):631–641.

[16] **Wan H**, HolmÃ©n A, NÃ¥gÃ¥rd M, Lindberg W. Rapid screening of pKa values of pharmaceuticals by pressure-assisted capillary electrophoresis combined with short-end injection. Journal of Chromatography A. 2002; 979(1):369 – 377. http://www.sciencedirect.com/science/article/pii/S0021967302012621, doi: https://doi.org/10.1016/S0021-9673(02)01262-1, 15th International Symposium on Microscale Separations and Analysis.

[17] **Cabot JM**, Fuguet E, Rosés M, Smejkal P, Breadmore MC. Novel Instrument for Automated p K_a Determination by Internal Standard Capillary Electrophoresis. Analytical Chemistry. 2015 Jun; 87(12):6165–6172. http://pubs.acs.org/doi/10.1021/acs.analchem.5b00845, doi: 10.1021/acs.analchem.5b00845.

[18] **Reijenga J**, van Hoof A, van Loon A, Teunissen B. Development of Methods for the Determination of pK_a Values. Analytical Chemistry Insights. 2013 Jan; 8:ACI.S12304. http://journals.sagepub.com/doi/10.4137/ACI.S12304, doi: 10.4137/ACI.S12304.

[19] **Sterling T**, Irwin JJ. ZINC 15 – Ligand Discovery for Everyone. Journal of Chemical Information and Modeling. 2015 Nov; 55(11):2324–2337. http://pubs.acs.org/doi/10.1021/acs.jcim.5b00559, doi: 10.1021/acs.jcim.5b00559.

[20] **Baell JB**, Holloway GA. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. Journal of Medicinal Chemistry. 2010 Apr; 53(7):2719–2740. http://pubs.acs.org/doi/abs/10.1021/jm901137j, doi: 10.1021/jm901137j.

[21] **Saubern S**, Guha R, Baell JB. KNIME Workflow to Assess PAINS Filters in SMARTS Format. Comparison of RDKit and Indigo Cheminformatics Libraries. Molecular Informatics. 2011 Oct; 30(10):847–850. http://doi.wiley.com/10.1002/minf.201100076, doi: 10.1002/minf.201100076.

[22] eMolecules Database Free Version;. Accessed: 2017-06-01. https://www.emolecules.com/info/products-data-downloads.html.

[23] OEChem Toolkit 2017.Feb.1;. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[24] **Shelley JC**, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. Journal of Computer-Aided Molecular Design. 2007 Dec; 21(12):681–691. http://link.springer.com/10.1007/s10822-007-9133-z, doi: 10.1007/s10822-007-9133-z.

[25] Schrödinger Release 2016-4: Epik Version 3.8;. Schrödinger, LLC, New York, NY, 2016.

[26] OEMolProp Toolkit 2017.Feb.1;. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[27] **Wishart DS**. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Research. 2006 Jan; 34(90001):D668–D672. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkj067, doi: 10.1093/nar/gkj067.

[28] **Pence HE**, Williams A. ChemSpider: An Online Chemical Information Resource. Journal of Chemical Education. 2010 Nov; 87(11):1123–1124. http://pubs.acs.org/doi/abs/10.1021/ed100697w, doi: 10.1021/ed100697w.

[29] NCI Open Database, August 2006 Release;. https://cactus.nci.nih.gov/download/nci/.

[30] Enhanced NCI Database Browser 2.2;. https://cactus.nci.nih.gov/ncidb2.2/.

[31] **Kim S**, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. PubChem Substance and Compound databases. Nucleic Acids Research. 2016 Jan; 44(D1):D1202–D1213. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv951, doi: 10.1093/nar/gkv951.

[32] NCI/CADD Chemical Identifier Resolver;. https://cactus.nci.nih.gov/chemical/structure.

[33] **Bemis GW**, Murcko MA. The properties of known drugs. 1. Molecular frameworks. Journal of medicinal chemistry. 1996; 39(15):2887–2893.

[34] OEMedChem Toolkit 2017.Feb.1;. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[35] Sirius T3 User Manual, v1.1. Sirius Analytical Instruments Ltd, East Sussex, UK; 2008.

[36] **Avdeef A**, Comer JEA, Thomson SJ. pH-Metric log P. 3. Glass electrode calibration in methanol-water, applied to pKa determination of water-insoluble substances. Analytical Chemistry. 1993; 65(1):42–49. https://doi.org/10.1021/ac00049a010, doi: 10.1021/ac00049a010.

[37] **Takács-Novák K**, Box KJ, Avdeef A. Potentiometric pKa determination of water-insoluble compounds: validation study in methanol/water mixtures. International Journal of Pharmaceutics. 1997; 151(2):235 – 248. http://www.sciencedirect.com/science/article/pii/S0378517397049077, doi: https://doi.org/10.1016/S0378-5173(97)04907-7.

[38] OEDepict Toolkit 2017.Feb.1;. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[39] **Fraczkiewicz R**. In Silico Prediction of Ionization. In: *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* Elsevier; 2013.http://linkinghub.elsevier.com/retrieve/pii/B978012409547202610X, doi: 10.1016/B978-0-12-409547-2.02610-X.

[40] **Manallack DT**. The pK(a) Distribution of Drugs: Application to Drug Discovery. Perspectives in Medicinal Chemistry. 2007 Sep; 1:25–38.